IRSTI 55.01.11

Ulan Alsiyeu[1*] ,*Zhasdauren Duisebekov*[2]
[1,2]SDU University, Kaskelen, Kazakhstan
*e-mail: 201107066@stu.sdu.edu.kz

A REGION-SPECIFIC APPROACH TO TRAFFIC SIGN RECOGNITION IN KAZAKHSTAN: A COMPARATIVE STUDY OF RESNET-101, MOBILENETV2, AND YOLOV8

Abstract. This research addresses the critical need for accurate traffic sign recognition in Kazakhstan, which is essential for enhancing road safety and developing advanced driver-assistance systems (ADAS). We created a comprehensive dataset tailored to Kazakhstan's traffic conditions and evaluated three state-of-the-art deep learning models: ResNet-101, MobileNetV2, and YOLOv8. Among these, YOLOv8 demonstrated superior performance, achieving 89.2% accuracy, 89.6% precision, 88.9% recall, and an 89.2% F1-score. This study highlights the effectiveness of tailored data augmentation techniques and the potential of YOLOv8 for real-time traffic sign recognition in dynamic environments, significantly contributing to the improvement of ADAS and road safety in Kazakhstan.

Keywords: Traffic Sign Recognition, Deep Learning, YOLOv8, ResNet-101, MobileNetV2, Data Augmentation, Advanced Driver Assistance System (ADAS)

Introduction

The increasing prevalence of advanced driver-assistance systems (ADAS) underscores the critical need for precise and reliable traffic sign recognition (TSR). Effective TSR is essential not only for enhancing road safety but also for ensuring the seamless operation of ADASin diverse and dynamic traffic environments. While significant advancements have been made in TSR using deep learning techniques, the unique traffic conditions and regulatory signage in specific regions often pose additional challenges that generic models may not adequately address. In Kazakhstan, the lack of region-specific datasets and tailored research on TSR has hindered the development of optimized ADAS solutions suited to local conditions. This research aims to bridge this gap by creating a comprehensive dataset specifically for traffic signs found in Kazakhstan and evaluating the performance of three state-of-the-art deep learning models: ResNet-101, MobileNetV2, and YOLOv8. The novelty of this study lies in its region-specific approach to dataset creation and model evaluation. By focusing on the unique characteristics of traffic signs in Kazakhstan, we aim to develop more accurate and reliable TSR systems that can be directly applied to enhance road safety in the region. Furthermore, the comparative analysis of ResNet-101, MobileNetV2, and YOLOv8 models provides insights into their respective strengths and weaknesses, guiding the selection of the most suitable model for practical implementation in Kazakhstan's road environments. Through this research, we aim to enhance the functionality of ADAS in Kazakhstan, ultimately contributing to improved road safety and the advancement of autonomous driving technologies in the region.

Literature review

Deep learning has revolutionized TSR by providing robust and scalable solutions for real-time detection and recognition of traffic signs. Convolutional Neural Networks (CNNs) have been extensively used due to their ability to learn hierarchical features from images. The YOLO (You Only Look Once) algorithm has emerged as a popular choice for real-time object detection, including TSR. YOLO's strength lies in its speed and precision, making it suitable for applications in ADAS. A systematic review by the authors [1] highlighted the widespread use of YOLO in TSR, emphasizing its application in various datasets and the sophisticated metrics used to evaluate its performance. The study identified common challenges in real-world implementations, such as varying lighting conditions and occlusions. Another notable work by Li and Huo [2] improved the YOLOv5 algorithm by integrating the Convolutional Block Attention Module (CBAM), enhancing the model's feature extraction capabilities and recognition accuracy, particularly for small and dense traffic signs. The study reported a 4.09% increase in mean Average Precision (mAP) compared to the standard YOLOv5 model. Moreover, Li and Wang [3] applied a CNN with the MobileNet architecture for TSR. MobileNet, designed for mobile and embedded vision applications, demonstrated efficient performance due to its lightweight structure. The use of batch normalization and advanced techniques significantly improved the recognition accuracy, showcasing the potential of these models in real-world scenarios. Lim et al. [4] explored the use of ensemble learning with CNNs, combining ResNet50, DenseNet121, and VGG16 models through majority voting. Their approach achieved remarkable accuracy rates on the GTSRB dataset, on the BTSD dataset, and on the TSRD dataset, demonstrating the robustness and efficacy of ensemble methods in TSR. The literature on TSR illustrates significant advancements through deep learning models, innovative data augmentation techniques, and the development of real-time systems. Models like YOLOv5, ResNet, and MobileNet, combined with robust data augmentation strategies, have shown remarkable performance.

Research methods

Data Collection

The dataset for this study was meticulously curated to ensure comprehensive coverage of traffic signs specific to a region in Kazakhstan. Our data collection process involved the following steps:

1. Legal Document Extraction: The primary source of our dataset was the official documentation provided by the legal framework of the Republic of Kazakhstan. This document, "Resolution of the Government of the Republic of Kazakhstan dated November 13, 2014, №1196," served as the foundational basis for our image dataset. We extracted visual representations of 220 unique instances of road traffic signs, covering six distinct classes, directly from this document.

2. Real-World Image Collection: To supplement the structured dataset, we collected 20 real-world images capturing various road scenarios using a car dashcam. These images were taken across different regions and environments in Kazakhstan, ensuring a diverse and realistic representation of the road environment as encountered in everyday situations. This approach provided contextual relevance to our dataset, which is crucial for training models to recognize traffic signs under realistic conditions.

3. Data Augmentation: Addressing class imbalance and instance scarcity is pivotal for developing robust traffic sign recognition models. We employed an extensive data augmentation process to enhance the diversity and realism of our dataset. Each traffic sign image underwent several transformations, including:

a. Geometric Transformations: Using the imgaug Python package, we applied five distinct geometric transformations such as rotation, shear, scaling, cropping, and translation to each sign image. These transformations simulate various perspectives and distortions that traffic signs might undergo in real-world conditions.

b. Color and Deformation Augmentations: To further increase the variability, we applied color filters and deformations, including brightness adjustments, noise addition, Gaussian blur, linear contrast, median blur, affine transformations, perspective transforms, and JPEG compression.

c. Background Integration: Leveraging the cv2 Python package, each augmented sign image was seamlessly integrated into the 20 real-world background images. This step was crucial for simulating the placement of traffic signs in diverse real-world contexts, ranging from urban streets to rural settings.

d. Environmental Conditions: We subjected each composite image to weather and lighting condition alterations using the albumentations package. Conditions applied include rain, snow, fog, sun flare, and variations in lighting (day, night, dawn effect), ensuring that the dataset reflects the challenges faced by advanced driving assistance systems in recognizing traffic signs under adverse conditions.

e. Obstacle-Based Augmentation: To further enhance the realism and robustness of the dataset, we introduced an obstacle-based augmentation technique. This involved overlaying obstacles such as vehicles, pedestrians, and trees on the traffic sign images, simulating real-world scenarios where signs might be partially obstructed

Figure 1. Augmented bike lane sign images

Through these rigorous data collection and augmentation processes, we created a comprehensive and representative dataset tailored to the unique traffic conditions and regulatory signs of Kazakhstan.

Models

To evaluate the effectiveness of traffic sign recognition systems for the region-specific dataset in Kazakhstan, we implemented and compared three state-of-the-art deep learning models: ResNet-101, MobileNetV2, and YOLOv8. Each model was chosen for its distinct architectural advantages and capabilities in handling various aspects of image recognition tasks.

1. ResNet-101:

a. Architecture: ResNet-101 is a deep convolutional neural network that employs a residual learning framework, which helps mitigate the vanishing gradient problem by allowing gradients to flow through shortcut connections. It consists of 101 layers, enabling it to learn complex features and patterns from the dataset.

b. Implementation: We implemented ResNet-101 using the PyTorch framework. The network was initialized with pre-trained weights from ImageNet, which provides a solid starting point for transfer learning. The final layers were adjusted to classify the specific traffic sign classes in our dataset.

c. Training: The model was trained using the Adam optimizer with an initial learning rate of 0.001. Data augmentation techniques were applied during training to improve the model's robustness and generalizability. The model was

trained for 50 epochs with a batch size of 32, using a cross-entropy loss function.
2. MobileNetV2:

a. Architecture: MobileNetV2 is a lightweight convolutional neural network designed for mobile and embedded vision applications. It uses depth wise separable convolutions to reduce the number of parameters and computational cost, making it efficient for real-time applications.

b. Implementation: MobileNetV2 was implemented using the TensorFlow framework with pre-trained weights from ImageNet. The network's final layers were modified to match the number of traffic sign classes in our dataset.

c. Training: The model was trained with the Adam optimizer and a learning rate of 0.0001. Data augmentation techniques, such as random cropping, rotation, and color adjustments, were utilized. Training was conducted for 50 epochs with a batch size of 32, optimizing for categorical cross-entropy loss. 3. YOLOv8: a. Architecture: YOLOv8 (You Only Look Once version 8) is an advanced object detection model that excels in real-time detection tasks. It frames object detection as a single regression problem, directly predicting bounding boxes and class probabilities from full images in one evaluation. b. Implementation: YOLOv8 was implemented using the Darknet framework. The model was initialized with pre-trained weights from the COCO dataset. The output layers were adapted to detect and classify the traffic signs in our dataset. c. Training: YOLOv8 was trained using stochastic gradient descent (SGD) with a learning rate of 0.001 and momentum of 0.9. Data augmentation included random scaling, cropping, and color distortions. The model was trained for 100 epochs with a batch size of 16, utilizing a mean squared error loss for bounding box prediction and a binary cross-entropy loss for classification.

Алсиеу Ұлан[1] *,Дүйсебеков Жасдаурен*[2]
[1,2]SDU University, Қаскелең, Қазақстан
*e-mail: 201107066@stu.sdu.edu.kz

ҚАЗАҚСТАНДАҒЫ ЖОЛ БЕЛГІЛЕРІН ТАНУДЫҢ АЙМАҚТЫҚ ТӘСІЛІ: RESNET-101, MOBILE NET V2 ЖӘНЕ YOLOV8 САЛЫСТЫРМАЛЫ ЗЕРТТЕУІ

Андатпа. Бұл зерттеу Қазақстан жол белгілерін дәл танудың маңыздылығын қарастырады, бұл жол қауіпсіздігін арттыру және жетілдірілген жүргізушіге көмек көрсету жүйелерін (ADAS) дамыту үшін өте маңызды. Біз Қазақстанның жол жағдайларына бейімделген жан-жақты деректер жиынтығын жасадық және ResNet-101, MobileNetV2 және YOLOv8 атты үш заманауи терең оқыту моделін бағаладық. Олардың ішінде YOLOv8 89,2% дәлдік, 89,6% нақтылық, 88,9% шақыру және 89,2% F1-есеп бойынша жоғары нәтижелер көрсетті. Бұл зерттеу нақты деректерді ұлғайту әдістерінің тиімділігін және YOLOv8-дің динамикалық ортада нақты уақыттағы жол белгілерін танудағы әлеуетін көрсетеді, бұл Қазақстандағы ADAS және жол қауіпсіздігін жақсартуға айтарлықтай үлес қосады.

Түйін сөздер: Жол белгілерін тану, Терең оқыту, YOLOv8, ResNet-101, MobileNetV2, Деректерді ұлғайту, Жетілдірілген жүргізушіге көмек көрсету жүйесі

Алсиеу Улан[1] *,Дүйсебеков Жасдаурен*[2]

[1,2]SDU University, Каскелен, Казахстан

*e-mail: 201107066@stu.sdu.edu.kz

РЕГИОНАЛЬНЫЙ ПОДХОД К РАСПОЗНАВАНИЮ ДОРОЖНЫХ ЗНАКОВ В КАЗАХСТАНЕ: СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ RESNET-101, MOBILE NET V2 И YOLOV8

Аннотация. Это исследование рассматривает критическую необходимость точного распознавания дорожных знаков в Казахстане, что необходимо для повышения безопасности дорожного движения и разработки передовых систем помощи водителю (ADAS). Мы создали комплексный набор данных, адаптированный к дорожным условиям Казахстана, и оценили три современных модели глубокого обучения: ResNet-101, MobileNetV2 и YOLOv8. Среди них YOLOv8 продемонстрировала наивысшую производительность, достигнув точности 89,2%, прецизионности 89,6%, полноты 88,9% и F1-меры 89,2%. Это исследование подчеркивает эффективность методов увеличения данных и потенциал YOLOv8 для распознавания дорожных знаков в реальном времени в динамичной среде, что значительно способствует улучшению ADAS и безопасности дорожного движения в Казахстане.

Ключевые слова: Распознавание дорожных знаков, Глубокое обучение, YOLOv8, ResNet-101, MobileNetV2, Увеличение данных, Система помощи водителю

Received 27 May 2024